

# (Paper's Draft) The Effects of an Aging Population on Military Recruitment and Retention: A Data-Driven Analysis

Lázaro Damasceno

January 9, 2026

## Abstract

Resumo em PT-BR.

## Abstract

Resumo em EN.

## 1 Introduction

## 2 Theoretical Framework

## 3 Methodology

### 3.1 Programming Language For Data Analysis

### 3.2 Datasets

The selected datasets were the *Predicted Populations of Brazil (2000-2070)*, provided by the Brazilian Census Office (BCO)<sup>1</sup>, and the *Military Service*, provided by the Brazilian Army (BA).<sup>2</sup>

The Government data is published on the Brazilian Government Open Data Portal. The portal was institutionalized by the Presidential Decree no. 8.777/2016.<sup>3</sup>

The military data includes 18 CSV files, whose total data volume is 4.9 GB. As Polars was slow to load them, Apache Spark was used. The total number of columns was 22. 15 were selected. The reduction of the number of columns was, more or less, 40.91%.

---

<sup>1</sup><https://www.ibge.gov.br/estatisticas/sociais/populacao/9109-projecao-da-populacao.html>

<sup>2</sup><https://dados.gov.br/dados/conjuntos-dados/servico-militar>

<sup>3</sup>[https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2016/decreto/d8777.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm)

The columns are:

The selected columns are:

The next step was turn the military data into parquet. 4.9 GB of data were reduced to X MB. X parquet files were created. They weight X MB. In other words, the GB data was reduced X%. As the new transformed data is only X% of the original data, Polars can handle it.

## **4 Discussion**

## **5 Results**

## **6 Final Remarks**